

Amendments to the Claims:

This listing of claims will replace all prior versions, and listings, of claims in the application:

Listing of Claims:

1. (currently amended) A method performed by a computer system, the method comprising:

retrieving, by a processor associated with the computer system, a first plurality of uniform resource locators (URLs), where one or more URLs of the first plurality of URLs include a parameter string comprising at least one parameter and a value associated with the at least one parameter;

selecting, by a processor associated with the computer system, one or more parameters present in at least a particular number of parameter strings of, respectively, the first plurality of URLs;

selecting, by a processor associated with the computer system, a first URL from the retrieved first plurality of URLs, where the first URL includes the selected one or more parameters;

generating, by a processor associated with the computer system, a second plurality of different URLs ~~having including, respectively,~~ different parameter combinations of the selected one or more parameters ~~selected parameters, where the parameter combinations include each combination of the selected one or more parameters;~~

retrieving, by a processor associated with the computer system, content using the first URL;

retrieving, by a processor associated with the computer system, contents ~~content~~ using, respectively each of the second plurality of different URLs;

comparing, by a processor associated with the computer system, the content retrieved using the first URL to the content retrieved using each of the second plurality of different URLs;

identifying, based on the comparing, one of the parameter combinations, that, when present in a particular URL, results in retrieving content that is approximately the same as the content corresponding to the first URL, the identifying being performed by a processor associated with the computer system; [[and]]

generating, by a processor associated with the computer system, one or more URL rewrite rules based on the identified one of the parameter combinations; and

indexing the first plurality of URLs and the second plurality of URLs based on the one or more URL rewrite rules.

2. (canceled)

3. (currently amended) The method of claim 1, further comprising:

~~performing the selecting a first URL, the generating a second plurality of different URLs, the retrieving content using the first URL, the retrieving content using the plurality of URLs, the comparing the content, and the identifying one of the parameter~~

combinations, for multiple different first URLs of the first plurality of URLs, each first URL including the one or more parameters; and

generating the one or more URL rewrite rules for the identified one of the parameter combinations for each of the first URLs.

4. (currently amended) The method of claim 3, where the rewrite rules specify that parameters₁ that do not occur in a threshold number of the identified one of the parameter combinations₂ are to be removed.

5. (previously presented) The method of claim 1, where each rewrite rule applies to a particular web site or web host.

6. (previously presented) The method of claim 1, where the identified one of the parameter combinations includes a minimum number of parameters with respect other ones of the parameter combinations.

7. (currently amended) A method, performed by a computer system, for converting a uniform resource locator (URL) into a canonical form of the URL, the method comprising:

receiving a URL that refers to content and that includes a parameter string including one or more parameters and values associated with the one or more parameters; selecting, by a processor of the computer system, a rewrite rule [[by]] including:

receiving a plurality of URLs that each includes a particular parameter string, where the particular parameter string includes a combination of the one or more parameters selected from the parameter string included in the received URL, and

identifying parameters₁ of the one or more parameters₁ that do not result in retrieving substantially different content, when present in a URL₁ including:

retrieving first content corresponding to a first URL₁ of the plurality URLs, the first URL including a first combination of parameters,

retrieving second contents corresponding, respectively, to a second plurality of URLs that include, respectively, different parameter combinations that include each combination of the first combination of parameters, where, for each of the second plurality of URLs, the first combination of parameters includes at least one parameter not included in the one of the second plurality of URLs, and

identifying one or more of the second plurality of URLs that is associated with second content, the second contents, that is not substantially different from the first content;

applying, by a processor of the computer system, the rewrite rule to the received URL by removing the parameters₁ that do not ~~contribute to~~ result in retrieving substantially different content from the received URL; and

outputting the rewritten URL as the canonical form of the URL.

8-10. (canceled)

11. (previously presented) The method of claim 7, where the rewrite rule applies to a particular web site or web host.

12. (currently amended) One or more devices comprising:
at least one fetch bot to download content on a network from locations specified by uniform resource locators (URLs);

a content manager to extract URLs from the downloaded content;

a rewrite component to

receive a URL that refers to content and that includes a parameter string including at least one parameter and a value associated with the at least one parameter,

apply a predetermined rewrite rule to the URL that removes the at least one parameter from the URL when removing the at least one parameter does not ~~affect~~ change the content referred to by the URL, where determining the predetermined rewrite rule ~~is determined by~~ includes:

receiving a plurality of URLs that include parameter strings comprising combinations of parameters and comprising at least one parameter and a value associated with the at least one parameter, and

identifying parameters in the parameter strings that do not result in retrieving substantially different content, when present in a URL[[:]],

including:

retrieving first content corresponding to a first URL, of the plurality URLs, the first URL including a first combination of parameters,

retrieving second content corresponding, respectively, to a second plurality of URLs that include, respectively, different parameter combinations that include each combination of the first combination of parameters, where, for each of the second plurality of URLs, the first combination of parameters includes at least one parameter not included in the one of the second plurality of URLs,
and

identifying one or more of the second plurality of URLs that are associated with second content that is not substantially different from the first content, and

output the rewritten URL as the canonical form of the URL; and
a URL manager to store the canonical form of the URL.

13-15. (canceled)

16. (previously presented) The one or more devices of claim 12, where each rewrite rule applies to a particular web site or web host.

17. (currently amended) A system comprising:
one or more devices comprising:

means for receiving a first uniform resource locator (URL) including a parameter string, where the parameter string includes one or more parameters and values associated with the one or more parameters;

means for retrieving content corresponding to the first URL;

means for retrieving content corresponding to a plurality of URLs ~~having~~ including, respectively, different parameter combinations of the one or more parameters, where the one or more parameters are selected from the parameter string, where the parameter combinations include each combination of the one or more parameters;

means for identifying the parameter combination from the plurality of URLs that corresponds to content that is approximately the same as the content corresponding to the first URL and that contains a minimum number of parameters compared to other parameter combinations; and

means for generating one or more URL rewrite rules based on the identified parameter combination.

18. (currently amended) A computer-readable memory device including programming instructions executed by a processor, the programming instructions comprising:

instructions for receiving a first uniform resource locator (URL) including a parameter string, where the parameter string includes one or more parameters and values associated with the one or more parameters;

instructions for retrieving content corresponding to the first URL;

instructions for retrieving content corresponding to a plurality of URLs ~~having~~
including, respectively, different parameter combinations of the one or more parameters,
where the one or more parameters are selected from the parameter string, where the
parameter combinations include each combination of the one or more parameters;

instructions for identifying the parameter combination from the plurality of URLs
that corresponds to content that is approximately the same as the content corresponding
to the first URL and that includes a minimum number of parameters; and

instructions for generating one or more URL rewrite rules based on the identified
parameter combination.

19. (previously presented) The system of claim 17, where the different parameter
combinations comprise an individual parameter of the one or more parameters, or a
combination of two or more parameters of the one or more parameters.

20. (currently amended) The computer-readable memory device of claim 18,
where the instructions for receiving a first URL, the instructions for retrieving content
corresponding to the first URL, the instructions for retrieving content corresponding to a
plurality of URLs, and the instructions for identifying the parameter combination are
performed for multiple first URLs, each first URL including the one or more parameters,
and where the one or more URL rewrite rules specify that parameters₁ that do not occur
in a threshold number of the identified parameter combinations₁ are to be removed.

21. (currently amended) The system of claim 17, further comprising:
means for determining whether the content_i that corresponds to the plurality of URLs_i is approximately the same as the content₁ that corresponds to the first URL₁ using a similarity hash function.

22. (previously presented) The computer-readable memory device of claim 18, where the rewrite rules specify that parameters that do not occur in a threshold number of the identified parameter combinations are to be removed.

23. (new) The method of claim 1,
where comparing the content retrieved using the first URL to the content retrieved using each of the second plurality of different URLs includes using a hash table to compare the content retrieved using the first URL to the content retrieved using each of the second plurality of different URLs, and
where the identified one of the parameter combinations, when present in the particular URL, results in retrieving content that differs from but is substantially similar to the content corresponding to the first URL.

24. (new) The method of claim 7,
where identifying one or more of the second plurality of URLs that is associated with second content, of the second contents, that is not substantially different from the first content includes comparing, using a hash table, the first content to each of the second contents, and

where the second content, associated with the identified one or more of the second plurality of URLs, differs from but is substantially similar to the first content retrieved using the first URL.

25. (new) The one or more devices of claim 12,
where the rewrite component, when identifying one or more of the second plurality of URLs that is associated with second content, of the second contents, that is not substantially different from the first content, compares, using a hash table, the first content to each of the second contents, and

where the second content, associated with the identified one or more of the second plurality of URLs, differs from but is substantially similar to the first content retrieved using the first URL.

26. (new) The system of claim 17,
where the identifying means is further for using a hashing table to identify content associated URLs, of the plurality of URLs, that differs from but is substantially similar to the content associated with the first URL.

27. (new) The computer-readable memory device of claim 18, where the instructions for identifying the parameter combination is further to compare, using a hashing table, the content that corresponds to the plurality of URLs to the content corresponding to the first URL, where the content corresponding to the plurality of URLs differs from but is substantially similar to the content corresponding to the first URL.